DOCUMENT RESUME

ED 160 646 TM 007 831

AUTHOR Burton, Nancy W.
TITLE Assessment As Exploratory Research: A Theoretical Overview.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
PUB DATE Mar 78
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Data Analysis; Educational Assessment; Educational Research; *Evaluation Methods; Experimenter Characteristics; *Models; National Surveys; Research; Research Design; Researchers; *Research Methodology; Scientific Methodology; *Scientific Research; *Social Science Research
IDENTIFIERS *Exploratory Research; Kuhn (T S); National Assessment of Educational Progress; *Paradigms

ABSTRACT

With the view that social scientists conduct too little exploratory research, the author outlines the methods and attitudes of the exploratory analyst and describes the National Assessment of Educational Progress (NAEP) as a source of research data. Exploratory analysis, in which no paradigm is used, is contrasted with confirmatory research, which is designed to extend a paradigm to the solution of a new problem. Three methods of organizing exploratory research are described: (1) ordinary-concept research, which is organized around concepts or catch words; (2) social problems research, which is organized around important social problems; and (3) mono-method research, which is organized around a particular analytic method. Skepticism about which results are considered true and which are in error, attention to outlying data, and an unwillingness to make assumptions are described as characteristics of the exploratory researcher. Since exploratory research is most often done by reanalyzing data collected for some other purpose, the breadth, precision and accuracy of the survey data available from the National Assessment of Educational Progress is described, with an emphasis on its usefulness to the exploratory researcher. Sample paradigms concerning the conceptual framework of human learning are appended. (JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

22.05

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Education Commission of the States

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM"

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED160646

ASSESSMENT AS EXPLORATORY RESEARCH:

A THEORETICAL OVERVIEW

DRAFT COPY

By Nancy W. Burton

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
1860 Lincoln Street, Suite 700
Denver, Colorado 80295

TM007 831

ASSESSMENT AS EXPLORATORY RESEARCH: A THEORETICAL OVERVIEW

by:
Nancy Burton

National Assessment of Educational Progress
March 1978

Presented at the 1978 Annual Meeting of the American Educational Research Association (Toronto) as part of a symposium entitled:

DATA EXPLORATION AS RESEARCH: THE NATIONAL ASSESSMENT

Chairman: Francis V. Corrigan
National Center for Education Statistics

Assessment as Exploratory Research: Theoretical Overview

This paper is about exploratory research. My major thesis is that social scientists generally, and educators particularly, do too little exploratory research. They tend to dismiss exploratory methods as "merely descriptive." However, I hold that descriptive, exploratory methods are precisely appropriate to areas of science, like education, that lack a guiding paradigm. Perhaps because educators see experimental methods, particularly since the introduction of modern statistics, as so powerful and so elegant; perhaps because they hold to a "bootstrap" view of scientific progress; or perhaps because until recently exploratory methods lacked a clear, modern exposition, educators have tended to avoid exploratory research.

In this paper, I hope to present exploratory research to the community of educational researchers for a fresh consideration. John Tukey and Frederic Mosteller have recently published two books -- Exploratory Data Analysis and Data Analysis and Regression -- which present the methods and, more important, embody the attitudes of the exploratory analyst. I will try to highlight these attitudes and then introduce a data base -- the National Assessment of Educational Progress -- which is particularly well-designed for data exploration. The rest of the papers in this symposium will present the National Assessment data in more detail and from a variety of points of view. We are hoping to encourage the research community to help us in the immense undertaking of exploring the National Assessment data.

Though much of the first part of my paper is devoted to defining exploratory research, I will begin with a brief introduction of the term. Most of the procedures we normally think of as scientific, such as

experimental studies or mathematical modeling work, depend on a paradigm.[1] The purpose of these studies is to extend the paradigm to the solution of a new problem and, in the process, to further confirm the paradigm. Kuhn calls this "normal science;" Tukey and Mosteller, in emphasizing the methodology of normal science, speak of "confirmatory research." I treat "normal science" and "confirmatory research" as synonymous but as distinct from "exploratory research." In exploratory analysis, no paradigm is assumed. In fact, one of the purposes of exploratory research is to find a paradigm.

It may be helpful to think of a continuum between exploratory and confirmatory poles. Exploration and confirmation do not differ either because in one we assume our theories to be <u>true</u> or because in the other we confront the raw data without any theories. Even in the most established science, the paradigm is not taken as true -- just as quite well-established. Even in the most exploratory area, the analyst does not confront the data completely without prejudice -- he or she has some hypotheses about the data. Rather, the continuum between exploratory and confirmatory analysis is a continuum of confidence in one's hypotheses. There is a clear line of demarcation on that continuum, because at some point the analyst ceases to explore alternative hypotheses and begins to confirm (or disconfirm) a single hypothesis.

Education is in a stage where no paradigm has been universally accepted. There is competition among explanations for actions: is a propositional organization sufficient to explain human learning? If it is, must we assume that the deep structure of language is unimportant? Is language a series of simple S-R bonds? This lack of a commonly-shared paradigm helps to explain the confusion in educational research noted by Gene Glass in his presidential

[1] See Appendix A for my definition of "paradigm."

address to the American Educational Research Association:

> In education, the findings are fragile; they vary in confusing irregularity across contexts, classes of subjects, and the countless other factors . . . This is particularly true of those questions that are more properly referred to as outcome evaluation than analytic research. Research on attention in learning or concept formation may sometimes progress along a systematic course; studies are designed from the findings of previous studies and researchers have a sense of what is known and what must be asked next. Outside the laboratory, however, inquiry is less cumulative and evolutionary (Glass, 1976, p. 3).

Some of the confusion in educational research results can be attributed to the complexity of education, or the lack of laboratory control. Much more, however, can be attributed to the lack of any central, guiding theory. Researchers tend to concentrate on concepts with similar names, such as "instructional media," but they share no paradigm and no common set of clearly-explicated relationships and laws, and so their results conflict.

The situation in education can be compared, for example, to very early stages of electrical research.

> The history of electrical research in the first half of the eighteenth century provides a more concrete and better known example of the way a science develops before it acquires its first universally received paradigm. During that period there were almost as many views about the nature of electricity as there were important electrical experimenters, men like Hauksbee, Gray, Desaguliers, Du Fay, Nollett, Watson, Franklin and others. All their numerous concepts of electricity had something in common -- they were partially derived from one or another version of the mechanico-corpuscular philosophy that guided all scientific research of the day. In addition, all were components of real scientific theories, of theories that had been drawn in part from experiment and observation and that partially determined the choice and interpretation of additional problems undertaken in research. Yet though all the experiments were electrical and though most of the experimenters read each other's work, their theories had no more than a family resemblance (Kuhn, 1970, pp. 13-14).

Science Without Paradigms

There are essentially two ways to practice science without a paradigm. One may carry on with confirmatory research, finding something other than

a paradigm around which to organize research; or, one can switch to exploratory research. In this section, I present examples of attempts to practice normal research without a paradigm.

Some scientists attempting to practice confirmatory research without a paradigm organize their work around ordinary concepts or catch words. A clear instance of the consequences of such ordinary-concept research can be seen in the conflicting results of much research in education. Consider "individualized instruction," for example. The various researchers do not agree on what individualized instruction is or, even if they do, they do not agree on the results they expect from it. Like other ordinary concepts, their ideas bear only a family resemblance. The necessary consequence is that, unless we are very lucky, their results conflict in unexplained ways, and thus the research does not have the cumulative growth of knowledge associated with a science.

An alternative to ordinary concept research is to organize research around important social problems. Unfortunately, these are often the most complex, confusing instances of reality -- the least likely to respond to a scientific solution. Scientific paradigms are more likely to start simple and only eventually be extended to fit complex phenomena. (Copernican astronomy predated a moon shot by over four hundred years.) It is revealing that many scientific revolutions are reducible to a simple image: Archimedes' bath, Newton's apple. The simple organizing image can easily generalize to other problems. On the other hand, organizing one's work around social problems practically guarantees that there will be none of the orderly development and progress usually

associated with a science. That is because the central problem is so complex that any solutions are almost bound to be ad hoc.

Still other scientists attempt to organize their research around some particular analytic method. I believe that the search for a rigorous way to organize scientific work has led to the rather common occurrence of "methodological evangalists" in the social sciences. These scientists build a career of carrying Rasch models or Bayesian statistics or Multiple Linear Regression to their colleagues.

I have discussed three ways of organizing research without a paradigm: ordinary-concept research, social problems research, and "monomethod" research. In each case, the alternative is unlikely to lead to the orderly advancement of knowledge. Why, then, do scientists persist in using confirmatory methods?

In the past, philosophers of science have seen the development of a science as a kind of bootstrap procedure in which fairly elaborate concepts become mini-theories which eventually coalesce into full-fledged paradigms. For example, both Kepler's and Galileo's work were seen as being subsumed by Newtonian physics and, subsequently, that Newtonian physics was embraced in the general theory of relativity. Under that view of science, one would attempt to practice normal research on whatever concepts or hunches were at hand, trusting the normal process of science continually to refine and expand the paradigms.

Thomas Kuhn, in The Structure of Scientific Revolutions, argued convincingly that science does not develop so smoothly. While some parts of earlier science have survived intact, others -- ptolemaic epicycles, phlogiston, the aether -- have been lost. These lost concepts were so important to the

earlier science that their loss, in Kuhn's word, amounts to a revolution. Kuhn envisions a period of crises as the earlier image is lost. This is a hiatus in which ordinary science is not practiced. A similar phenomenon occurs as each student learns the controlling paradigm in his area of science. Because Kuhn's episodic view of science explains such historical inconveniences as phlogiston, and is also an intuitively appealing explanation of one's personal learning of science, many scientists now accept his explanation.[2]

Even though they reject the bootstrap view of science, many practitioners in non-paradigmatic areas still use confirmatory, normal-research methods. For one thing, the kind of dispute about fundamental assumptions natural to pre-paradigmatic areas seems undignified when compared to the community of purpose natural to practitioners under a paradigm. The disputants are unfairly compared to the beyond-the-fringe characters, like Lysenko, who for non-scientific reasons refuse to accept the prevailing paradigm. But, in pre-paradigmatic areas, fundamental assumptions must be disputed -- to lack a paradigm means to lack safe assumptions.

Just as the skeptical attitude of the exploratory researcher may seem undignified, his analysis methods seem crude next to the elegant models possible under a paradigm. In paradigmatic research, the model defines what is real and what is error; in exploratory research, the peculiar, deviant data points may be the very ones that lead to a new paradigm.

In statistics, for example, the methods most appropriate to data exploration are usually called "descriptive statistics." Introductory texts

---

[2]For example, in a recent two-issue treatment of progress in science in Daedalus, nearly every article referred, at least obliquely, to Kuhn.

usually illustrate these methods with such non-scientific applications as baseball averages. Exploration is seldom discussed. Similarly, historians or philosophers of science seldom dwell on the exploratory phases of science. Kuhn says as much as any in the following passage:

> In the absence of a paradigm or some candidate for paradigm, all of the facts that could possibly pertain to the development of a given science are likely to seem equally relevant. As a result, early fact-gathering is a far more nearly random activity than the one that subsequent scientific development makes familiar. Furthermore, in the absence of a reason for seeking some particular form of more recondite information, early fact-gathering is usually restricted to the wealth of data that lie ready to hand. The resulting pool of facts contains those accessible to casual observation and experiment together with some of the more esoteric data retrievable from established crafts like medicine, calendar making and metallurgy. Because the crafts are one readily accessible source of facts that could not have been casually discovered, technology has often played a vital role in the emergence of new sciences.
>
> But thought this sort of fact-collecting has been essential to the origin of many significant sciences, anyone who examines, for example, Pliny's encyclopedic writings or the Baconian natural histories of the seventeenth century will discover that it produces a morass. One somehow hesitates to call the literature that results scientific. The Baconian "histories" of heat, color, wind, mining, and so on, are filled with information, some of it recondite. But they juxtapose facts that will later prove revealing (e.g., heating by mixture) with others (e.g., the warmth of dung heaps) that will for some time remain too complex to be integrated with theory at all.* In addition, since any description must be partial, the typical natural history often omits from its immensely circumstantial accounts just those details that later scientists will find sources of important illumination. Almost none of the early "histories" of electricity, for example, mention that chaff, attracted to a rubbed glass rod, bounces off again. That effect

*Compare the sketch for a natural history of heat in Bacon's Novum Organum, Vol. VIII of The Works of Francis Bacon, ed. J. Speeding, R. L. Ellis, and D. D. Heath (New York, 1869), pp. 179-203.

seemed mechanical, not electrical.** Moreover, since the casual fact-gatherer seldom possesses the time or the tools to be critical, the natural histories often juxtapose descriptions like the above with others, say, heating by antiperistasis (or by cooling), that we are now quite unable to confirm.*** Only very occasionally, as in the cases of ancient statics, dynamics, and geometrical optics, do facts collected with so little guidance from pre-established theory speak with sufficient clarity to permit the emergence of a first paradigm.

This is the situation that creates the schools characteristic of the early stages of a science's development. No natural history can be interpreted in the absence of at least some implicit body of intertwined theoretical and methodological belief that permits selection, evaluation and criticism. If that body of belief is not already implicit in the collection of facts -- in which case more than "mere facts" are at hand -- it must be externally supplied, perhaps by a current metaphysic, by another science, or by personal and historical accident. No wonder, then, that in the early stages of the development of any science different men confronting the same particular phenomena, describe and interpret them in different ways. What is surprising, and perhaps also unique in its degree to the fields we call science, is that such initial divergences should ever largely disappear.

**Duane Roller and Duane H. D. Roller, The Development of the Concept of Electrical Charge: Electricity from the Greekes to Coulomb ("Harvard Case Histories in Experimental Science," Case 8; Cambridge, Massachusetts, 1954), pp. 14, 22, 28, 43. Only after the work recorded in the last of these citations do repulsive effects gain general recognition as unequivocally electrical.

***Bacon, op. cit., pp. 235, 337, says, "Water slightly warm is more easily frozen quite cold." For a partial account of the earlier history of this strange observation, see Marshall Clagett, Giovanni Marliani and Late Medieval Physics (New York, 1941), Chapter IV.

(Kuhn, 1970, pp. 15-17).

Exploratory Research and Exploratory Researchers

The central principle of exploratory research is that there is no way to tell truth from error. In confirmatory research, the model (derived from the paradigm) defines "truth" as the data that fits the model and "error" as the residual data. So long as the paradigm works, the residuals can

safely be ignored. In exploratory research, we can neither treat the data so cavalierly nor the model so confidently.

In exploratory research, one still deals with both models and data, but the emphasis on each and the researchers' attitudes toward each are different. Both exploratory and confirmatory researchers -- simply because they are human beings -- work from within what contemporary epistemologists call the "conceptual framework."[3] The important point about the conceptual framework is that no human beings experience simple, raw, unstructured facts. They always have some concept or hypotheses or model or point of view about their experience.

So, exploratory researchers use models to analyze the data; no more than other people can researchers step outside their conceptual framework. However, researchers can impose a strict skepticism about their assumptions. This skepticism explains why exploratory researchers constantly try to make their assumptions explicit and, whenever possible, attempt to vary their assumptions to look at the data from several points of view.

It is difficult to be skeptical about assumptions. The major function of the conceptual framework is to allow people to make assumptions about reality so they can go ahead and act. In particular, the function of normal scientific procedures is to make the most of a set of assumptions. Perhaps because it is so unusual to treat assumptions skeptically, Kuhn calls exploratory research extraordinary science.

Techniques exist to help analysts vary their assumptions, though they may need to go so far afield as psychiatry or literature to find those techniques discussed. For example, the literary techniques of

[3]See Appendix A for a more complete discussion of the conceptual framework.

analogy or metaphor can be used in the attempt to find alternate assumptions: what if the growth of a city were like the growth of a tree? Even when specific analogues to other fields do not suggest themselves, simply shifting the borders of the field of inquiry may help. The trend toward hyphenated disciplines in the social sciences -- such as anthropological linguistics -- examplifies this exploratory motive.

Reorganizing at different levels of generality is another way of changing focus. Some thinkers see this specific change of focus as crucial to scientific advance:

> For every discipline in its early stages of development there exists an antidiscipline. For many-body physics, particle physics; for chemsitry, many-body physics; for molecular biology, chemistry; for cellular biology, molecular biology; and so forth. With the word antidiscipline, I wish to emphasize the special adversary relation that exists initially between the studies of adjacent levels of organization. This relationship is also creative, and with the passage of a great deal of time it becomes fully complementary.
>
> In this article I will argue that biology has now moved close enough to the social sciences to become their antidiscipline. . .
>
> If it is true that biology is the antidiscipline of the social sciences, the past failure of the social sciences to develop a common body of theory is understandable. The reason is that the relevant branches of biology -- neurobiology and sociobiology -- are only now becoming mature enough to attain a juncture with the social sciences. Although it would be premature to say that biology can revolutionize the social sciences, at least to the extent that chemistry has revolutionized biology, it would be equally premature to say that it cannot (Wilson, 1977, p. 127).

Even if the reduction of social science into biology is not as complete as Wilson predicts, such a reorganization may still provide insight. Suppose the field of inquiry is human learning. Reducing the inquiry to the level of cells may show that some areas of potential study may be explained genetically. Or, going one step higher in generality to observe group

social behavior may also shed light on the mechanisms controlling individual learning.

These are three rather formal ways of varying one's assumptions. Many other possible techniques -- from using different mathematical models on the same data to playing word-association games -- can be imagined and deserve systematic treatment. However, now I will turn to a brief discussion of the data attitudes characteristic of exploratory research.

Though exploratory researchers tend to treat their models or concepts with more skepticism than do confirmatory researchers, they treat their data with more reverence. Peculiar data might be accidental, but, on the other hand, their explanation might lead to a paradigm. Recall Kuhn's example of chaff that bounces away from a glass rod -- this was dismissed as a mere mechanical reaction until the process of static electricity was understood.

In reviewing Tukey's Exploratory Data Analysis (1977), Howard Wainer points out:

> He describes the basic equation of data analysis as
>
> DATA = FIT + RESIDUAL
>
> and uses this throughout as a unifying concept (with minor variants, as in the discussion of smoothing techniques where he uses
>
> DATA = SMOOTH + ROUGH).
>
> Professor Tukey's interest in looking at residuals is in evidence throughout. Using the basic equation he shows that frequently one wants to separate the main part (the preliminary fit) of the data from the whole by subtraction and look at the residuals. After separating the data into two parts, Tukey very carefully states that the data points with large residuals are "set aside," never discarded. He emphasizes that although such points may be considered contaminants, sometimes the contaminants are what are of interest (in South Africa, good building stone is "contaminated" by diamonds) (Wainer, 1977, pp. 635-636).

In fact, after fitting a model to the "smooth," the analyst may fit another model to the "rough" and see what the combination of the two fits tells about the data.

Another Tukey technique is a series of data displays and summaries that expose outlying data. The outlying data points are often individually named and described. For example, in Exhibit 1 there are two displays of the heights of the highest points in fifty states. Exhibit 1A is the "stem and leaf"[4] which contains all the raw data; Exhibit 1B is a summary "box-and-whisker"[5] plot of the same data. The box-and-whisker plot, in particular, separates the outliers from the main body of data so both may be considered independently.

Since Tukey does not discard outliers, he selects summary statistics which are quite resistant to the influence of outliers. Thus, he normally uses such order statistics as medians and hinges (the first and third quartiles). While it is fair to take account of odd data, they should not change the view of the entire data set.

I have mentioned two data-related behaviors of the exploratory analyst -- agnosticism with regard to which part of the data should be considered "true" and which "error" and attention to outlying data. A third characteristic -- the unwillingness to make assumptions -- is closely related to the last. Because of this trait, the analyst might tend to

[4]The "stem," to the left of the line, gives the first two digits of the number and the leaves -- to the right of the line -- are the third digit. Thus, the highest point in Alaska is 203 hundred (20,300) feet. There are five states whose highest point is less than 1,000 feet, and their highest points are 300, 400, 500, 800 and 800 feet.

[5]The "box" contains all data between the first and third quartiles ("hinges"); the middle line is the median. The "whiskers" extend to cover the entire range of the data.

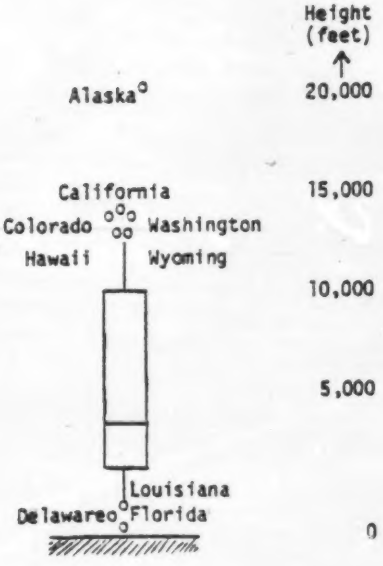
Exhibit 1

The Heights of the Highest Points in each State[6]

A) STEM-and-LEAF -- Unit 100 Feet

0*	34688	Fla, Del, La, Miss, RI
1	236788	
2	003448	
3	24556	
4*	01489	
5	03347	
6	367	
7	2	S. Dak
8*	8	Texas
9		
10		
11	2	Oregon
12*	678	Idaho, Ariz, Mont
13	12588	Nev, N. Mex, Utah, Wyo, Hawaii
14	445	Wash, Colo, Calif
15		
16*		
17		
18		
19		
20*	3	Alaska

B) BOX-and-WHISKER



[6]Tukey (1977).

retain outlying data, even if they never led to new insights. The outlying data would prevent the analyst from making glib assumptions or thoughtless interpretations. The analyst would always have to ask how such-and-such a computation will work -- given these peculiar data. In fact, the unwillingness to make data-related assumptions is merely a specific manifestation of the pre-paradigmatic scientist's skepticism about concepts and assumptions.

Both confirmatory and exploratory research employ concepts and data, but in different ways. The two also differ in direction. Confirmatory research is deductive; it goes from general principals to more and more complex specific applications. In exploratory, one starts at the data and attempts to move toward generalization. One attempts, if possible, to move upward in generality by discarding assumptions and discarding details -- but tentatively, since one has no paradigm to separate the wheat from the chaff.

The purpose of exploratory data analysis is to allow the data to lead one to a paradigm. Those who discuss scientific discovery emphasize the "Eureka!" aspect -- the fact that important discoveries very often come as sudden inspirations or, quite often, in dreams. I would like to emphasize that the discoveries are also based on a great deal of experience with data. The new model of how the world works may be inspired, but it is based on experience. The experiential data may be one's personal experience (the displacement of water in the bathtub), data collected under some previous theory (the retrograde movements of the stars), or data collected "randomly" (medieval beast books).

So far in this paper, I have presented the idea of a paradigm and attempted to establish that a paradigm is fundamental to science as an orderly, growing body of knowledge. Following Kuhn, I claimed a radical

distinction between a science with a paradigm and those scientific areas which have never had paradigms or in which the paradigm no longer functions. There is a parallel and equally important distinction between the confirmatory methods used in paradigmatic areas and the exploratory methods used to search for a paradigm. Finally, I tried to define exploratory research, by contrasting it with confirmatory research and by showing the attitudes and behaviors of a model exploratory reseacher.

In the final sections of this paper, I will discuss some of the attributes of data desirable for exploratory research and present a specific example of a desirable data base.

The Data

Any kind of data may help one form a new paradigm. In an already-established science, the very anomalous evidence that leads to crisis often becomes the central data upon which the new paradigm is based. This evidence is initially difficult to interpret because the data are both generated and described in the terms of the old paradigm. For example, both Priestley and Lavoiser "discovered" oxygen. However, Priestley persisted in seeing it as depholgisticated air, while Lavoiser, who was already dissatisfied with the phlogiston theory, was able to perceive oxygen in terms of "a combustion theory which led to a reformulation of chemistry so vast that is usually called the chemical revolution" (Kuhn, 1970, p. 56).

In general, exploratory research is most often done by reanalyzing data collected for some other purpose. In education, modern technology provides a wealth of data that can be used in an exploratory manner. there are some rather narrow areas -- particularly in psychology -- where paradigms seem tc exist and so educationally-relevant experimental data

are available on such topics as concept formation, memory and reading comprehension. These data can be combined and reanalyzed to provide insight into educational issues.

There is a growing number of well-executed comparative evaluations. Even without a theory, it is possible sometimes to define a set of competing treatments and to evaluate their relative merits.[7]

There are the many "ordinary concept" and "social problems" studies of the kind routinely produced by professors and graduate students in many departments.

Finally, there are several kinds of data not collected as a part of an experiment or a comparative randomized field trial, but rather collected because of bureaucratic requirements. Floods of information are generated yearly through the evaluation requirements of the various titles of the Elementary and Secondary Education Act. Huge amounts of similar data have been collected by private textbook and test publishers. Education agencies at all governmental levels routinely collect head-count information about facilities, enrollment, attendance, finance and so on. Particularly, survey data on achievement has been collected in many cities, states and since 1969 for the nation as a whole.

These data are not all equally useful for exploration. The data from well-designed experiments or comparative evaluations are likely to be so specifically focused that a secondary use is only a rare accident.

---

[7] Perhaps the clearest examples of this kind of data can be seen in the medical field -- comparisons of drugs such as The National Halothane Study (Bunker et. al, 1969) or psychiatric therapies (Smith and Glass, 1977). For a general discussion of such randomized field trials, see Gilbert, Mosteller and Light (1977).

The mess of studies I have called "ordinary concept" or "social problems" research, as well as the many program evaluation studies, suffer under the double problems of a rather specific primary purpose and poorly-explicated assumptions (because of the lack of a paradigm).

In fact, the simple head-count information has some special virtues for exploratory analysis. In the first place, census studies lack the "advocate bias" present in all the other studies -- which are nearly all done to demonstrate the virtue of the researcher's method. (The advocacy is just as present in the best-designed studies: the main difference is that in the good studies, justice keeps her blindfold on.) In the second place, the head-count data are often collected from many different locales; in the best instances, from an entire population or a scientific sample. And finally, the data, since they are routinely required, tend to be both simple (so neither respondents nor analysts can misinterpret them too badly) and publicly available.

In the next sections, I would like to discuss the particular virtues of the survey data available from the National Assessment of Educational Progress.

The National Assessment Design

The assessment was designed to provide a periodic survey of achievement in several learning areas. Each year, a probability sample of students at three age levels (and most years, a sample of adults 26 to 35 years old) is tested in one or two learning areas. Since the first assessment in 1969, baseline information has been collected in career and occupational development, music, art and literature; writing, reading,

and citizenship/social studies have been reassessed once; and science and mathematics have been reassessed twice.

I will discuss three important elements of the National Assessment design: breadth, precision and accuracy. These three elements are the ones that make the data particularly useful in the exploration of diverse hypotheses that can lead to the formation of paradigms.

Breadth

The assessment items are written to fit a very broad objectives structure. Recently, NAEP objectives have been set up as matrices which cross Bloom-and-Krathwohlesque cognitive and affective response categories[8] (rows) with content-defined columns. These content categories are usually a compromise between various views of what is being taught, what soon will be taught, and what ought to be taught. The result of all these compromises is usually that the content columns are more comprehensive, and less logically elegant, than an individual effort would have been. The matrix provides a very broad, taxonomic structure for a learning area.

Each assessment can be seen as a long-term periodic sweep over a learning area. The method of development prevents the assessment content from being very focused or very doctrinaire. The assessment procedures have been designed assuming that the direction of education over

[8] Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. Benjamin Bloom, editor; McKay Co., Inc., New York, 1972; Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain. D. R. Krathwohl, B. S. Bloom, R. B. Masia, editors; David McKay Co., Inc., New York, 1972.

the next fifteen years is unpredictable and that a broad, systematic search pattern is most likely to pick up unpredictable changes.[9]

The assessment covers broad content categories and as many categories of cognitive and affective response as possible. There are also other dimensions on which the assessment self-consciously attempts breadth. Objective, multiple-choice items are used whenever appropriate, but attitude and self-report exercises are also included, as are individual interviews, performances (such as writing, drawing or playing a musical instrument) and manipulation exercises (estimating the weight of objects, demonstrating a scientific principle by using a model, or demonstrating the most efficient way of collating and stapling a newsletter). An attempt is made to include enough easy items to find out what everybody can do. (Not what everybody "should" do.)

Variety is also introduced in assessment by the addition of new items. Up to one-half of the exercises are published after every assessment and not reused. The major motivation for this practice is the attempt to make assessment results clear and understandable; a major consequence is that a large number of new items, written to measure revised objectives, are introduced every few years.

Various methods of scoring and analysis are applied to the same data, which allow the results to be understood from many perspectives. For example, an essay may be rated on a scale of inventiveness and then

[9]For areas in which interest can be predicted, the assessment has instituted several small-scale probes which are designed for fast development, assessment and reporting. Probe topics have included "Basic Life Skills" (non-3-R skills) at age 17 and energy knowledge and attitudes among young adults.

scored descriptively for mechanical errors, grammatical sophistication, vocabulary, or stylistic strategies. Results of the set of exercises are also summarized in various ways. The essay results may be included in writing content summaries, or summaries over several basic skills areas or in a summary of attitudes expressed by 17-year-olds (as for the essay topic "A Woman's Place is in the Home.") The results can be reported raw or adjusted for mechanical ability or attitudes toward writing. The performance of black females on essays may be contrasted with their performance on practical items.

These attempts to broaden the assessment assure that many different kinds of change will be detectable. The breadth of the assessment also assures that any one change can not be described in detail. The early NAEP reports in particular appeared to be made up of isolated bits of information: drops in the bucket. In later reports, attempts have been made to organize the information in various kinds of after-the-fact themes, to make the results more general. These thematic reports signaled the start of a trend toward more and more diverse analysis and reporting. Though the analyses multiply, however, only occasionally will they be sufficient to overcome the general rule that the breadth required to detect unpredictable changes is incompatible with the depth required to explain them.

Precision and Accuracy

The next important assessment design attributes, after breadth of coverage, are precision and accuracy of estimation. If one tries to

measure unpredictable changes by spreading a very broad net, every impingement on that net must be sensitively detected.

By far, the most important contributor to the assessment's accuracy is its sample design. The norming samples for important commercial tests -- renormed every ten to twenty years -- compare poorly with the yearly assessment sample. The assessment uses a scientifically-designed probability sample, and implements it with utmost care. Letters are sent to states, districts and schools; district meetings are held; NAEP employees select the samples of students and administer the tests. As a result, final cooperation rates -- counting all district, school and student noncooperation -- average above ninety percent for ages 9 and 13 and at about eighty-five percent for 17-year-olds by dint of scheduling make-up sessions and follow-up sessions and pursuing dropouts and early graduates to their homes. The completeness of the final sample determines the possible amount of bias in national assessment results.

Many other efforts are made to reduce bias. Each item is reviewed by as many as twenty people to detect cultural or technical defects that would bias results for some group of respondents. The "I don't know" option is included on most items to reduce the biasing effect of guessing. Formats and directions are made as simple as possible to reduce difficulties unrelated to the task being measured. For example, an audio tape is used to reduce reading difficulties; answers are recorded in test booklets rather than on separate answer sheets; complex tasks are administered to one respondent at a time. All of these efforts are made to increase the accuracy of NAEP measures: so that the raw results of a single task can be interpreted simply, without caveats.

The precision of NAEP results depends largely on the reduction of sampling errors. Precision is separate from accuracy: the classic example of the distinction is a scale that measures very precisely, within ounces, but adds a bias of five pounds to everything it weighs. Precision is generally determined by sample size: the more schools and the more students sampled, in general, the smaller the sampling errors. Sampling errors are also related to the amount of "clustering." Thus, a more precise sample is possible for a school if students are sampled from many classrooms than if entire classrooms are sampled. Analogously, clustering is also reduced if each student takes very few items or, at least, if the items he receives are unrelated. Both forms of clustering introduce correlations which reduce the amount of unique information in the data collected, and, therefore, reduce the effective sample size. The NAEP sample is designed so that the "standard error" of an exercise of medium difficulty is one to two percent; the standard error of easy or difficult exercises will be smaller.

Analysis of Assessment Data

Assessment is primarily designed to describe educational trends and to locate problems and successes. The task of explaining educational trends -- that is, of confirming the causes of those trends -- is definitely beyond the scope of assessment. Confirming causes is the proper task of controlled experimentation. However, between simple description and causal explanation there is room for an important set of secondary purposes for assessment.

For many social movements, survey data provide the only possible controlled observations. These are the broad social movements -- like

the settlement of the West or the rise of the Frisbee -- that are impossible to manipulate experimentally because they are within no one's control. For example, the large-scale court-ordered desegregation of schools in the south was preceded and followed by the first two national science assessments.[10] The wide-scale introduction of electronic calculators was antedated by the first assessment of mathematics. Assessment data are perhaps our only chance to understand the effects of such social changes.

In other words, even though assessment cannot provide final explanations of causes, it <u>can</u> provide hypotheses about causes. For these broad-scale events, assessment data are the best-controlled source of causal information. Interpretation of assessment data in these areas can appear perilously close to an attempt at causal explanation. In all areas, however, hypothesis generation should be seen as distinct from hypothesis testing and enormously important in its own right.

In normal science, the paradigm generates hypotheses almost automatically, and hypotheses suggested by the data can be treated as serendipitous extras. In pre-paradigmatic areas, however, sources of hypotheses are scarce: they come from ordinary concepts, personal experiences, complex, problem-oriented evaluations and experiments from related disciplines. While all of these sources are useful, it should be seen that assessment data are likely to be much more useful, since they are so broad and since they are so carefully collected.

---

[10] In 1969, approximately seventy percent of southeastern blacks attended schools in which less than ten percent of the student body was white. By 1973, only about twenty percent of southeastern blacks attended such radically segregated schools (NAEP, 1976).

The business of pre-paradigmatic science is to generate and explore many hypotheses from many points of view. This kind of catholic questioning is most likely to uncover a fruitful new paradigm. Just as I have described assessment as a systematic scan of the facts, pre-paradigmatic science can be described as a systematic scan of concepts in search of a paradigm, where concepts are different combinations or structures of the facts.

In exploring assessment data, the analyst in essence is trying different organizations or structures for the results. The analyst can eliminate many implausible hypotheses by exploring the data. Thus, assessment can be seen as a proving ground in which many ideas can be proposed and eliminated and the few strong ideas can be located for further, more controlled, exploration. Since educators tend to practice normal science, few such exolorations have been attempted. To practitioners of normal science, assessment data seem crude and inappropriate. If I am right, however, in considering education to be pre-paradigmatic, exploration of assessment is a most promising way to begin to provide a systematic body of educational knowledge.

APPENDIX A

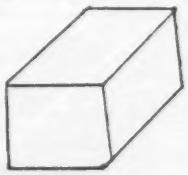
Paradigms

The Conceptual Framework

In analyzing how people learn, epistemologists note that our ideas are not like discrete marbles jostling about in our heads. Ideas are connected; they form a model or conceptual framework or point of view of reality. Otto Neurath likens this conceptual framework to a boat. One is afloat in it and thus can change it only slowly. One can change one's conception of reality only if there is a replacement; otherwise, the boat sinks.

Ludwig Wittgenstein uses a series of perceptual examples from Gestalt experiments to show how our beliefs about causes and relationships -- our "conceptual framework" -- become an integral part not only of our thinking about reality, but even a part of our perception of reality.

"You could imagine the illustration



appearing in several places in a book, a text-book for instance. In the relevant text something different is in question every time: here a glass cube, there an inverted open box, there a wire frame of that shape, there three boards forming a solid angle. Each time the text supplies the interpretation of the illustration.

But we can also see the illustration now as one thing now as another -- So we interpret it, and see it as we interpret it."

(Wittgenstein, 1953, p. 193e)

And, later:

> I suddenly see the solution of a puzzle-picture. Before, there were branches there; now there is a human shape. My visual impression has changed and now I recognize that it has not only shape and colour but also a quite particular 'organization'. -- My visual impression has changed . . . .
>
> (Wittgenstein, 1953, p. 196e)

Wittgenstein calls this sudden change in one's perception of an object the dawning of an aspect. It is particularly well-illustrated by the following example

which may be seen as either a duck or a rabbit. More easily than in the case of the jigsaw puzzle piece, one can experience the change in one's perception of the duck-rabbit.

> I see two pictures, with the duck-rabbit surrounded by rabbits in one, by ducks in the other. I do not notice that they are the same. Does it follow that I see something different in the two cases?
>
> (Wittgenstein, 1953, p. 195e)

One can physically feel the change in perception when one sees the duck-rabbit first as one, then as the other. This physical feeling is objectively-observable evidence for the effect of one's expectations on the raw data of the senses. As Wittgenstein remarks later,

> The colour of the visual impression corresponds to the colour of the object (this blotting paper looks pink to me, and is pink) -- the shape of the visual impression to the shape of the object (it looks rectangular to me, and is rectangular) -- but what I perceive in the dawning of an aspect is not a property of the object, but an internal relation between it and other objects.
>
> (Wittgenstein, 1953, p. 212e)

Seeing and thinking cannot be disentangled. A dog and a man may both see a pile of rocks, but a man usually also sees it as moraine or a cairn or a rock quarry or a pyramid -- that is, his perception also includes the idea of actions, agents, relations, causes -- and the concept is the glass through which he sees the object. Thus, our most basic contact with the world is not uncontaminated. It is filtered through the conceptual framework. This conceptual framework provides a model of the world and it allows us to form expectations (make predictions) about the objects and events in the world. Some of these images of the world are fruitful -- they allow us to build such highly-structured systems of relationships as those contained in successful scientific theories. On the other hand, the most basic axioms of a theory, no matter how rigorous the rest of the theory may be, are articles of belief about the world -- or, more precisely, metaphors or models or points of view about the world -- not fundamentally different from the assumptions in some not-so-well developed sector of the conceptual framework.

From Ordinary Concepts to Paradigms

Kuhn, like Wittgenstein, used the Gestalt phenomenon of the dawning of an aspect to explain what happens when a paradigm is learned.

> Instead, as we have already seen, normal science ultimately leads only to the recognition of anomalies and to crises. And these are terminated, not by deliberation and interpretation, but by a relatively sudden and unstructured event like the gestalt switch. Scientists then often speak of the "scales falling from the eyes" or of the "lightning flash" that "inundates" a previously obscure puzzle, enabling its components to be seen in a new way that for the first time permits its solution. On other occasions the relevant illumination comes in sleep.
>
> (Kuhn, 1970, pp. 122-123)

As I understand Kuhn's usage, "paradigm" is an honorific title for a special kind of concept. That is, what we normally call science is organized around especially good concepts. First, those concepts are rich or graphic. They allow people to generate many expectations about observable phenomena. These expecations provide the raw material for the normal business of science: to test expectations through carefully-designed experiments.

Second, a paradigm must allow for theoretical development. "Paradigm" and "theory" are closely-related terms. The concept of "paradigm" is vague -- it is associated with such words as image and expectations; in contrast, a theory is not vague -- it is the mathematical or logical formulation of the paradigm. This formalization transforms that which I have loosely called the "expectations" associated with a paradigm into explicitly-testable predictions. It formalizes a general view of a relationship into a specific law, like $E = MC^2$. While a paradigm must allow for theoretical development, it is clear that in practice the paradigm comes first. First, the earth is seen as part of a solar system; later, the exact model is worked out.

Finally, a paradigm must be accepted by a significant portion of a scientific community. This is necessary simply because "paradigm" is a word of social approbation, so that an unrecognized paradigm is just not, in fact, a paradigm. To be accepted by the scientific community, a paradigm must provide the scientist with work. Therefore, it must be rich in

testable expectations and must allow for theoretical development.* The third requirement is that the scientific community must be ready for the paradigm. So long as the old paradigm is generating normal scientific work, a new concept is not likely to gain acceptance. For example, the Ptolemaic concept of the earth-centered universe was a good paradigm. It lasted thousands of years. It could tolerate the many inelegant adjustments necessary to explain the hesitations, loops and backtracking of heavenly bodies from the earthly perspective a long time before the less intuitively-appealing sun-centered astronomy of Copernicus could emerge.

A Word About Method

In the preceding section, I defined a paradigm as a particularly rich kind of concept -- one with many consequences; and a theory as the operationalization of a paradigm -- that which generates specifically-testable hypotheses. In this section, I want to discuss the distinction between a paradigm and a method.

Paradigms are concrete. They are specific ideas about how some part of reality works: the earth goes around the sun. Methods are abstract. They are rules for working that may apply to more than one area. For example, the method of creating a methematical model for the behavior of objects applies not only to astronomy but also to economics and particle physics. However, there has been a great tendency in social science research to confuse methods with paradigms. For example:

*Some philosophers of science argue that if a paradigm does either, it is enough. Certainly, many social scientists are kept occupied by fitting models to data that are not precise enough to predict future behavior. (Or, at least to predict future behavior correctly). The question, really, is whether they are doing normal science or pre-paradigmatic science. I have opted to call their work pre-paradigmatic.

> There are some good indications that educational research may have reached a crisis stage with regard to its major Fisherian experimental design tradition. and perhaps that the paradigm has never worked (Magoon, 1977).

It is confusing to say that Fisherian experimental design is a "paradigm" that no longer works. Fisherian experimental design is a set of methods; specifically, it is a set of methods designed to test the hypotheses generated by a paradigm. Fisherian experimental design is in the tradition of normal research: the theory generates an hypothesis which the experiment then tests. Every separate experiment is cumulative in that each tends to confirm, to modify, or to disconfirm the theory. The problem with applying confirmatory methods to education is that we generally do not have educational theories to confirm.

REFERENCES

Bloom, Benjamin (ed.). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. New York: McKay Co., Inc., 1972.

Bunker J., Forrest W., Mosteller, F. and Vandom, L. The National Halothane Study: A Study of the Possible Association Between Halothane Anesthesia and Postsperative Hepatio Necrosis. Washington, D.C.: Superintendent of Documents, U.S. Government Printing Office, 1969.

Gilbert, J., Mosteller, F. and Light L. "Assessing Social Innovations: An Empirical Base for Policy," in Statistics and Public Policy. Edited by W. Fairley and F. Mosteller. Reading, Massachusetts: Addison-Wesley, 1977.

Glass, G. "Primary, Secondary and Methodology Analysis of Research," Educational Researcher, 5 (November, 1976), 3-8.

Krathwohl, D. R., Bloom, B. S., Masia, R. B. (eds.). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain. New York: McKay Co., Inc., 1972.

Kuhn, Thomas S. The Structure of Scientific Revolutions, Vol. II, No. 2, second edition, of International Encyclopedia of Unified Science. Chicago: University of Chicago Press, 1970.

Magoon, A. "Constructivist Research," Review of Educational Research, 47 (1977), 651-693.

Mosteller, F. and Tukey, J. Data Analysis and Regression. Reading, Massachusetts: Addison-Wesley, 1977.

NAEP. Science Achievement: Racial and Regional Trends, 1969-1973, Report No. BRS-1. Denver, Colorado: National Assessment of Educational Progress, 1976.

Smith, M. and Glass, G. "Meta Analysis of Psycho-Therapy Outcome Studies," American Psychologist, 32 (September, 1977), 752-760.

Tukey, J. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley, 1977.

Wainer, H. Review of J. W. Tukey's "Exploratory Data Analysis," Psychometrica, 42 (1977), 635-638.

Wilson, E. "Biology and the Social Sciences," Daedalus, 106, No. 4 (1977), 127-140.

Wittgenstein, Ludwig. Philosophical Investigations. New York: The Macmillan Company, 1953.